

CLAIMS

1. One or more computer-readable media having stored thereon a plurality of instructions that, when executed by one or more processors of a computer, cause the one or more processors to:

generate a bandwidth value indicating an average bandwidth used by a server to perform previous server operations;

receive a request for the server to perform an additional server operation;

compare the bandwidth value to a threshold; and

restrict the request in a first manner if the bandwidth value exceeds the threshold.

2. One or more computer-readable media as recited in claim 1, wherein the plurality of instructions further cause the one or more processors to:

store data indicative of a data transmission rate for each of the previous server operations; and

generate the bandwidth value by averaging the stored data transmission rates.

3. One or more computer-readable media as recited in claim 1, wherein the plurality of instructions cause the one or more processors to restrict the request in the first manner if both the bandwidth value exceeds the threshold and the additional server operation is of a first type.

1 4. One or more computer-readable media as recited in claim 3, wherein
2 the first type comprises a read operation.

3
4 5. One or more computer-readable media as recited in claim 3, wherein
5 the first type comprises a write operation.

6
7 6. One or more computer-readable media as recited in claim 3, wherein
8 the first type comprises a transmit operation.

9
10 7. One or more computer-readable media as recited in claim 1, wherein
11 the previous server operations and the additional server operation include one or
12 more of read operations, write operations, and transmit operations.

13
14 8. One or more computer-readable media as recited in claim 1, wherein
15 the plurality of instructions further cause the one or more processors to:
16 compare the bandwidth value to another threshold; and
17 restrict the request in a second manner if the bandwidth value exceeds the
18 other threshold.

19
20 9. One or more computer-readable media as recited in claim 8, wherein:
21 the threshold is less than the other threshold;
22 restricting the request in the first manner comprises delaying the request;
23 and
24 restricting the request in the second manner comprises blocking the request.
25

1 **10.** One or more computer-readable media as recited in claim 1, wherein
2 the restricting the request in the first manner comprises delaying the request.

3
4 **11.** One or more computer-readable media as recited in claim 1, wherein
5 the restricting the request in the first manner comprises blocking the request.

6
7 **12.** One or more computer-readable media as recited in claim 1, wherein
8 the restricting the request in the first manner comprises preventing the request
9 from being presented to the server.

10
11 **13.** One or more computer-readable media as recited in claim 1, wherein
12 the restricting the request in the first manner comprises preventing the request
13 from entering a request queue of the server.

14
15 **14.** One or more computer-readable media as recited in claim 1, wherein
16 the plurality of instructions include instructions that cause the one or more
17 processors to generate the bandwidth value by:

18 generating a value for each of the previous server operations by,

19 identifying a time interval duration between a start time of the
20 operation and the end time of the operation,

21 identifying a number of bytes transferred for the operation, and

22 dividing the number of bytes by the time interval duration; and

23 dividing a sum of the values of the previous server operations by the
24 number of previous server operations.
25

1 **15.** A method comprising:
2 receiving a request to perform a server operation; and
3 restricting performance of the request based at least in part on an average
4 bandwidth used in performing previous server operations.

5
6 **16.** A method as recited in claim 15, further comprising:
7 storing data indicative of a data transmission rate for each of the previous
8 server operations; and
9 generating the average bandwidth by averaging the stored data transmission
10 rates.

11
12 **17.** A method as recited in claim 15, wherein the restricting comprises
13 restricting the request if both the bandwidth value exceeds a threshold and the
14 server operation is of a first type.

15
16 **18.** A method as recited in claim 17, wherein the first type comprises a
17 read operation.

18
19 **19.** A method as recited in claim 17, wherein the first type comprises a
20 write operation.

21
22 **20.** A method as recited in claim 17, wherein the first type comprises a
23 transmit operation.

1 **21.** A method as recited in claim 15, wherein the restricting comprises:
2 comparing the average bandwidth to at least one of a first threshold and a
3 second threshold;
4 restricting the request in a first manner if the average bandwidth exceeds
5 the first threshold but does not exceed the second threshold; and
6 restricting the request in a second manner if the average bandwidth exceeds
7 the second threshold.

8
9 **22.** A method as recited in claim 21, wherein restricting the request in a
10 first manner comprises delaying the request and wherein restricting the request in
11 a second manner comprises blocking the request.

12
13 **23.** One or more computer-readable memories comprising computer-
14 readable instructions that, when executed by a processor, direct a computer system
15 to perform the method as recited in claim 15.

16
17 **24.** A host system comprising:
18 at least one network server;
19 an asynchronous thread queue to receive a request, from a client process, to
20 be performed by one of the at least one network server; and
21 a bandwidth throttling system, coupled to the asynchronous thread queue,
22 to determine whether performance of the request by the network server is to be
23 restricted based at least in part on an average bandwidth used by the network
24 server in performing previous operations.

1 **25.** A host system as recited in claim 24, wherein the bandwidth
2 throttling system is further to:

3 store data indicative of a data transmission rate for each of the previous
4 server operations; and
5 generate the average bandwidth by averaging the stored data transmission
6 rates.

7
8 **26.** A host system as recited in claim 24, further comprising an ancillary
9 function driver, coupled to the asynchronous thread queue, to couple the host
10 system to a network.

11
12 **27.** A host system as recited in claim 24, wherein the at least one
13 network server comprises a plurality of network servers, and wherein the
14 bandwidth throttling system determines whether performance of a request by a
15 particular one of the plurality of network servers is to be restricted based on the
16 average bandwidth used by that particular network server in performing previous
17 operations and independent of the average bandwidth used by other network
18 servers of the plurality of network servers in performing previous operations.

19
20 **28.** A host system as recited in claim 24, wherein the bandwidth
21 throttling system comprises:

22 a measurement subsystem to compute the average bandwidth used by the
23 network server in performing previous operations; and
24
25

1 a control subsystem, coupled to the measurement subsystem, to make the
2 determination and communicate how the asynchronous thread queue is to restrict
3 performance of the request.
4

5 **29.** A method comprising:
6 generating a value indicating a bandwidth used by a server to perform
7 previous server operations;
8 receiving a request for the server to perform an additional server operation;
9 and
10 restricting the request in a first manner if the value exceeds a first threshold
11 but does not exceed a second threshold, and restricting the request in a second
12 manner if the value exceeds the second threshold.
13

14 **30.** A method as recited in claim 29, further comprising:
15 storing data indicative of a data transmission rate for each of the previous
16 server operations; and
17 generating the value by averaging the stored data transmission rates.
18

19 **31.** A method as recited in claim 29, further comprising:
20 generating another value indicating bandwidth used by another server to
21 perform other previous server operations;
22 receiving another request for another server to perform another additional
23 server operation; and
24
25

1 restricting the other request in the first manner if the value exceeds a third
2 threshold but does not exceed a fourth threshold, and restricting the other request
3 in the second manner if the value exceeds the fourth threshold.
4

5 **32.** A method as recited in claim 29, wherein the first threshold is
6 different than the third threshold, and wherein the second threshold is different
7 than the fourth threshold.
8

9 **33.** A method as recited in claim 29, wherein the first manner comprises
10 delaying the request.
11

12 **34.** A method as recited in claim 29, wherein the second manner
13 comprises blocking the request.
14

15 **35.** A method as recited in claim 29, wherein the restricting the request
16 in the first manner comprises preventing the request from being presented to the
17 server.
18

19 **36.** A method as recited in claim 29, wherein the restricting the request
20 in the first manner comprises preventing the request from entering a request queue
21 of the server.
22

23 **37.** A method as recited in claim 29, wherein the value indicates an
24 average bandwidth used by the server.
25

1 **38.** A method as recited in claim 29, wherein the restricting comprises
2 restricting the request in the first manner or the second manner only if the request
3 is of a first type.

4
5 **39.** A method as recited in claim 38, wherein the first type comprises a
6 read operation.

7
8 **40.** A method as recited in claim 38, wherein the first type comprises a
9 write operation.

10
11 **41.** A method as recited in claim 38, wherein the first type comprises a
12 transmit operation.

13
14 **42.** One or more computer-readable memories comprising computer-
15 readable instructions that, when executed by a processor, direct a computer system
16 to perform the method as recited in claim 29.

17
18 **43.** One or more computer-readable media having stored thereon a
19 plurality of instructions that, when executed by one or more processors of a
20 computer, cause the one or more processors to:

21 generate a value indicating a bandwidth used by a server to perform
22 previous server operations;

23 receive a request for the server to perform an additional server operation;

24 compare the value to at least one of a first threshold and a second threshold;

25 and

1 delay the request if the value exceeds the first threshold but does not exceed
2 the second threshold, and block the request if the value exceeds the second
3 threshold.
4

5 **44.** One or more computer-readable media as recited in claim 43,
6 wherein the plurality of instructions further cause the one or more processors to:
7 store data indicative of a data transmission rate for each of the previous
8 server operations; and
9 generate the value by averaging the stored data transmission rates.
10

11 **45.** One or more computer-readable media as recited in claim 43,
12 wherein the value indicates an average bandwidth used by the server.
13

14 **46.** One or more computer-readable media as recited in claim 43,
15 wherein the delaying comprises delaying the request only if the request is of a first
16 type, and wherein the blocking comprises blocking the request only if the request
17 is of a first type.
18

19 **47.** One or more computer-readable media as recited in claim 46,
20 wherein the first type comprises a read operation.
21

22 **48.** One or more computer-readable media as recited in claim 46,
23 wherein the first type comprises a write operation.
24
25

1 **49.** One or more computer-readable media as recited in claim 46,
2 wherein the first type comprises a transmit operation.

3
4 **50.** A host system comprising:
5 at least one network server;
6 an asynchronous thread queue to receive a request, from a client process, to
7 be performed by one of the at least one network server; and
8 a bandwidth throttling system, coupled to the asynchronous thread queue,
9 to compare a value indicating a bandwidth used by the network server to perform
10 previous server operations to at least one of a first threshold and a second
11 threshold, to restrict the request in a first manner if the value exceeds the first
12 threshold but does not exceed the second threshold, and to restrict the request in a
13 second manner if the value exceeds the second threshold.

14
15 **51.** A host system as recited in claim 50, wherein the bandwidth
16 throttling system is further to:

17 store data indicative of a data transmission rate for each of the previous
18 server operations; and

19 generate the value by averaging the stored data transmission rates.
20
21
22
23
24
25

1 **52.** A host system as recited in claim 50, further comprising a plurality
2 of network servers, and wherein the bandwidth throttling system determines
3 whether to restrict a request for a particular one of the plurality of network servers
4 based on the value indicating bandwidth used by that particular network server in
5 performing previous operations and independent of the bandwidth used by other
6 network servers of the plurality of network servers in performing previous
7 operations.

8
9 **53.** A host system as recited in claim 50, wherein the bandwidth
10 throttling system comprises:

11 a measurement subsystem to compute, as the value, an average bandwidth
12 used by the network server in performing previous operations; and

13 a control subsystem, coupled to the measurement subsystem, to make the
14 determination and communicate how the asynchronous thread queue is to restrict
15 performance of the request.

16
17 **54.** One or more computer-readable media having stored thereon a
18 plurality of instructions that, when executed by one or more processors of a
19 computer, cause the one or more processors to:

20 determine a presently used bandwidth for each of at least one network
21 server by way of a data transmission rate measurement during execution of an
22 operation for each of said at least one network server that includes:

23 storing data indicative of said data transmission rate measurement of
24 said operation for n last most recently executed operations for each of said
25

1 at least one network server, wherein n is a positive integer greater than 2,
2 and

3 generating data indicative of an effective presently used bandwidth
4 for each of said at least one network server wherein said effective presently
5 used bandwidth includes an average of said data transmission rate
6 measurement over said n last most recently executed operations;

7 effect provision of a plurality of classes of service provided by one of said
8 at least one network server in a first manner, in response to said effective presently
9 used bandwidth for said one of said at least one network server that exceeds a first
10 threshold; and

11 effect provision of said plurality of classes of service provided by said one
12 of said at least one network server in a second manner that differs from said first
13 manner, in response to said effective presently used bandwidth for said one of said
14 at least one network server that exceeds a second threshold that is greater than said
15 first threshold.

16
17 **55.** One or more computer-readable media having stored thereon a
18 plurality of instructions that, when executed by one or more processors of a
19 computer, cause the one or more processors to:

20 store data indicative of a predetermined allocated data transmission
21 bandwidth for each of a plurality of network servers, each of said network servers
22 providing a plurality of classes of service;

23 determine an effective bandwidth for each of said network servers,
24 comprising:
25

1 calculating a bandwidth for each operation performed by each of
2 said network servers,

3 tabulating a count of bandwidth used by each of said network servers
4 in each of a plurality of last time intervals, and

5 averaging said tabulated count of bandwidth to obtain a value
6 indicative of said effective bandwidth;

7 delay, in response to said effective bandwidth for a one of said network
8 servers substantially corresponding to said predetermined allocated data
9 transmission bandwidth for said one network server, for delaying a first subset of
10 said plurality of classes of service provided by said one network server; and

11 reject, in response to said effective bandwidth for said one network server
12 exceeding said predetermined allocated data transmission bandwidth for said one
13 network server, requests for said first subset of classes of service provided by said
14 one network server and for delaying service for a second subset of said plurality of
15 classes of service provided by said one network server.

16
17 **56.** One or more computer-readable media having stored thereon a
18 plurality of instructions that, when executed by one or more processors of a
19 computer, cause the one or more processors to:

20 determine a presently used bandwidth for each of a plurality of network
21 servers by way of a data transmission rate measurement taken during execution of
22 an operation for each of said plurality of network servers, said determining
23 including:

24 storing data indicative of a predetermined allocated data
25 transmission bandwidth for each of said plurality of network servers, and

1 storing data indicative of a first threshold, wherein said first
2 threshold includes an indication of a differential from said predetermined
3 allocated data transmission bandwidth for each of said plurality of network
4 servers; and

5 effect provision of a plurality of classes of service provided by a first one of
6 said plurality of network servers in a manner that is individually defined for a
7 plurality of successively greater thresholds in response to said effective presently
8 used bandwidth of said first one of said plurality of network servers exceeding one
9 of said plurality of successively greater thresholds, wherein said manner defined
10 for said first one of said plurality of network servers differs from every other
11 manner defined for other ones of said plurality of network servers.
12
13
14
15
16
17
18
19
20
21
22
23
24
25